PB83-169722

Characterization of Heavy Residual Fuel Oils and Asphalts by Infrared Spectrophotometry Using Statistical Discriminant Function Analysis

(U.S.) Environmental Monitoring and Support Lab.
Cincinnati, OH

1974

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before comple

1. REPORT NO.	2.	3
EPA-600/J-74-001	JOURNAL ARTICLE	PB83-169722

4. TITLE AND SUBTITLE	5. REPORT DATE
Journal Article: Characterization of Heavy Residual Fuel Oils and Asphalts by Infrared Spectrophotometry Using Statistical Discriminant Function Analysis	1974
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) F. K. Kawahara, J. F. Santner, and E. C. Julian	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS Environmental Monitoring and Support Laboratory U.S. Environmental Protection Agency 26 W. St. Clair Cincinnati, OH 45268	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring and Support Laboratory U.S. Environmental Protection Agency 26 W. St. Clair Cincinnati, OH 45268	13. TYPE OF REPORT AND PERIOD COVERED Journal article, Feb. 1974
	14. SPONSORING AGENCY CODE EPA/600/06

15. SUPPLEMENTARY NOTES
REF: Analytical Chemistry, Vol. 46, No. 2, February 1974, pp. 266-273

16. ABSTRACT
Spilled asphaltic materials and heavy residual fuel oils, because of their high molecular weights, complexity, and physical nature, cannot be readily identified to a source since these materials are not usually amenable to analysis by gas chromatography with flame ionization detector. The limiting difficulty is the inability to vaporize and to separate the high molecular weight components in the GC column. Infrared spectrophotometry is a useful technique for characterization of these materials. A combination of infrared spectrophotometry, data treatment, data transformation, and discriminant function analysis through computer assistance has resulted in effecting a more precise and accurate method of distinguishing between these two heavy petroleum products. An established statistical technique has been successfully applied to a recently developed infrared procedure for heavier petroleum products to provide a useful and powerful technique for classification of samples.

Continued on pg. 266

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS (This Report)	21. NO. OF PAGES
Release to Public	Unclassified	9
	20. SECURITY CLASS (This page)	22. PRICE
	Unclassified	

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U.S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

U.S. Government is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

EPA-600/J-74-001
JOURNAL ARTICLE

Reprinted from ANALYTICAL CHEMISTRY, Vol. 46, Page 266, February 1974
Copyright 1974 by the American Chemical Society and reprinted by permission of the copyright owner

Characterization of Heavy Residual Fuel Oils and Asphalts by Infrared Spectrophotometry Using Statistical Discriminant Function Analysis

F. K. Kawahara, J. F. Santner, and E. C. Julian

U.S. Environmental Protection Agency, National Environmental Research Center, Analytical Quality Control Laboratory, Cincinnati, Ohio 45268

Spilled asphaltic materials and heavy residual fuel oils, because of their high molecular weights, complexity, and physical nature, cannot be readily identified to a source since these materials are not usually amenable to analysis by gas chromatography with flame ionization detector. The limiting difficulty is the inability to vaporize and to separate the high molecular weight components in the GC column. Infrared spectrophotometry is a useful technique for characterization of these materials. A combination of infrared spectrophotometry, data treatment, data transformation, and discriminant function analysis through computer assistance has resulted in effecting a more precise and accurate method of distinguishing between these two heavy petroleum products. An established statistical technique has been successfully applied to a recently developed infrared procedure for heavier petroleum products to provide a useful and powerful technique for classification of samples.

Among the many efforts for upgrading the environment are the control of stack emissions, the wise use of pesticides, reduction of pollutants in automotive exhausts, efficient management of solid wastes, and the improvement of water quality. One of the serious and continuing water problems is the discharge of oils into surface waters. Thus, analytical procedures for the identification of discharged petroleum products and the successful application of these techniques are significant contributions to law enforcement and environmental improvement.

Asphaltic materials and heavy residual fuel oils, because of their high molecular weights, complexity, and physical nature, cannot be readily identified to a source since these materials are not usually amenable to analysis by gas chromatography with flame ionization detector. The limiting difficulty is the inability to vaporize and to separate the high molecular weight components in the GC column. However, one feasible approach leading to identification is the analysis of the passively labeled fluorinated phenolic ethers and thioether derivatives of the weak acids present in discharged heavier petroleum products (*1*, *2*) by the use of electron capture detector gas chromatography.

Infrared spectrophotometry is a useful technique presented for the characterization of these heavier residual fuel oils. Initial screening tests will eliminate aromatic lube oils and No. 2 fuel oils. Differences in density, solubility in hexane, and much weaker infrared absorptivities distinguish these lighter products from the heavier products. Also, No. 1 and No. 2 fuel oils have distinctive infrared absorptivities at other wavenumbers.

An application of this technique is described by Kawahara and Ballinger (*3*). Numerous unidentified oil samples collected from the surface waters of several locations in the United States were analyzed by infrared spectrophotometry using peaks at six wavenumbers. For rapid field evaluation, the graphic method with the use of two key ratios, 810 cm^{-1}/1375 cm^{-1} and 810 cm^{-1}/720 cm^{-1}, was useful for the initial classification of unknown oil spill samples among asphalts, No. 6 fuel oils, No. 5 fuel oils, and a lube oil.

A demonstration of the effectiveness of the infrared method of ratios for heavier residual fuel oils was illustrated in the recent publication (*1*) where two different spills in the same general area resulted in samples being collected at two possible sources. The two different spills were characterized as being entirely unlike while a source sample was coupled to each spill and shown to be alike, respectively. In addition, each of the six ratios of infrared absorbances of commercial asphalts, No. 6 fuel oils, No. 5 fuel oils, and cutter stocks were distinctly characteristic for each product. The ratios of absorbance generated in that study are shown in Table I and are illustrative of the data used in this paper.

To further improve the characterization potential using infrared spectrophotometry and to eliminate the small indeterminant region previously observed, linear discriminant function analysis using 42 variables was applied to data points obtained previously (*3*).

(1) F. K. Kawahara, *J. Chromatogr. Sci.*, **10**, 629 (1972).

(2) F. K. Kawahara, Division of Water, Air and Waste Chemistry, 161st National Meeting, American Chemical Society, Los Angeles, Calif., March 28–April 2, 1971, Abstract No. 3.

(3) F. K. Kawahara and D. G. Ballinger, *Ind. Eng. Chem., Prod. Res. Develop.*, **9**, 553 (1970).

LINEAR DISCRIMINANT FUNCTION

A linear discriminant function is a function that is linear in the variables and discriminates or classifies data into one of the groups under consideration. (*4*). Two groups or products are considered in this paper—namely, heavy residual fuel oils and asphalts.

A separate linear discriminant function is obtained for each product. The forms of these two functions are:

$$z_a = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_5x_5 + \ldots + b_{36}x_{36} \quad (1)$$

$$z_o = b_0' + b_1'x_1 + b_2'x_2 + b_3'x_3 + b_4'x_4 + b_5'x_5 + \ldots + b_{36}'x_{36} \quad (2)$$

where z_a and z_o designate the asphalt and heavy residual fuel oil functions, respectively. The x_i's are the variables used in the problem. The coefficients or parameters (primed and unprimed b's) are determined by maximizing the ratio of the "between" group variance (heavy residual fuel oils and asphalts) to the "within" group variance (heavy residual fuel oils or asphalts). To evaluate both discriminant functions, the variable values of an unknown sample are used. The unknown would then be assigned to that product whose discriminant value, z_a or z_o, was a maximum or the larger. In addition to classifying a sample, the probability of correct classification into either group can be calculated. An example of both procedures with a specific example is given in a later section.

The linear discriminant function procedures as used in this paper will classify any sample as either a heavy residual fuel oil or an asphalt (*5, 6*). In the present form, it is a dichotomous classification procedure; however, the extension to any number of groups may be considered and is limited only by computer capability.

EXPERIMENTAL

For analyses of infrared spectra, samples were prepared by techniques described earlier (*3*). The thickness of the liquid cell or film was adjusted so that the transmittance readings varied from 80 to 95% of the full scale at 2925 cm^{-1}.

PREPARATION OF DATA FOR DISCRIMINANT ANALYSIS AND SUMMARY OF RESULTS

The data preparation prior to performing discriminant analysis is described below. The data for this problem in discriminant analysis consist of data elements of two groups (asphalts and heavy residual fuel oils), and of 69 and 77 observations, respectively. Each observation of data consists of measurements from seven peaks from which net absorbances are calculated. Peak and base-line distances for each peak were measured in centimeters to the nearest 0.01 cm. In order that the effects of variation in cell path length be minimized, 42 ratios of these absorbances were selected for calculation. Cells of fixed path length are now used where samples are sufficiently liquid. A schematic layout of absorbance ratios from data previously obtained (*3*) is shown in Table II.

In Table II, three replicates (R_{2111}, R_{2112}, and R_{2113}) are shown for Group II, heavy residual fuel oils, the first ratio of absorbances, 720 cm^{-1}/810 cm^{-1}; and sample number one. The notations for the three replicates, R_{2111}, R_{2112}, and R_{2113} are derived from the general notation, R_{ijkl} which stands for the original ratio of infrared absorbances.

The term i = 1 (asphalt), 2 (No. 6 fuel oil); j = 1, 2 ... 42 (ratio identification); k = 1, 2 ... 18 or 20 (sample number identification); and l = 1, 2, 3, 4, or 5 (non-uniform replicate identification).

Two changes in the data handling marked the difference between the previous analysis (*3*) and the analysis in this paper. In the previous paper, averages, not replicates, were used in the analysis. For example, $\bar{R}_{211} = (R_{2111} + R_{2112} + R_{2113})/3$ was used instead of the three replicates. In this paper, the three replicates R_{2111}, R_{2112}, and R_{2113} were used as such.

The second change involved was the standardization of the data of Table II before performing discriminant analysis. Data were standardized by subtracting from each table entry the corresponding sample column mean $\bar{R}_j$ which equals the sum of all ratios of absorbances in one column of Table II, including all samples and replicates within both groups, divided by n, (n = 146), the total number of replicates. This difference is divided by the sample column standard deviation, s_j over both groups. The term $s_j = \sqrt{\Sigma(R_j - \bar{R}_j)^2/(n-1)}$ where R_j = value of a single ratio, $\bar{R}_j$ = mean value of ratios of absorbances of the corresponding column, and n = number of replicates.

The total number of asphalt and oil samples and their respective replicates are indicated in Table II. The equation for the standardization of the ratio variable is

$$x_{ijkl} = (R_{ijkl} - \bar{R}_j)/s_j$$

where $\bar{R}$ and s_j values are given in Table III.

Discriminant analysis on the standardized data was performed with a computer program described by Dixon (*7*).

The output of the program (*7*), in part, consists of:

(1) At each step:
  (a) Variables included and F to remove.
  (b) Variables not included and F to enter.
(2) After the last step:
  (a) Classification functions (discriminant functions).
(3) For each case (observation):
  (a) The posterior probability of the case coming from each group.

Table I. Ratios of Infrared Absorbances of Enforcement Samples

Ratios of infrared absorbance	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
720 cm^{-1} / 1375 cm^{-1}	0.08	0.11	0.08	0.12	0.06
3050 cm^{-1} / 2925 cm^{-1}	0.16	0.07	0.16	0.09	0.14
1600 cm^{-1} / 1375 cm^{-1}	0.63	0.24	0.73	0.31	0.60
1600 cm^{-1} / 720 cm^{-1}	8.20	2.24	9.66	2.46	9.41
810 cm^{-1} / 1375 cm^{-1}	0.77	0.25	0.90	0.31	0.74
810 cm^{-1} / 720 cm^{-1}	9.96	2.29	11.89	2.53	11.60

(4) M. G. Kendall and A. Stuart, "The Advanced Theory of Statistics," Volume III, Hafner Publishing Company, New York, N.Y., 1946, Chapter 44.

(5) E. C. Julian, "Application of Discriminant Function Analysis to Infrared Data," Newsletter, No. 14, p 8, July (1972), J. B. Anderson, Ed., published by the Analytical Quality Control Laboratory, NERC, Environmental Protection Agency, 1014 Broadway, Cincinnati, Ohio 45202.

(6) J. F. Santner, "Status of Research Report," memorandum to W. J. Benoit, National Environmental Protection Agency, Cincinnati, Environmental Protection Agency, July 10, 1972.

(7) W. J. Dixon, "BMD, Biomedical Computer Programs," University of California Press, Berkeley, Calif., 1971, p 214a–214t.

Table II. Schematic Layout of the Absorbance Ratio Data

	No. of samples per group	No. of replicates per sample	Ratios of absorbances: $R_{i,1,k,l}$ 720 cm^{-1} / 810 cm^{-1}	$R_{i,1,k,l}$ 720 cm^{-1} / 870 cm^{-1}	...	$R_{i,42,k,l}$ 1600 cm^{-1} / 1460 cm^{-1}
Group I Asphalts $i = 1$	$k = 1$	$l = 1,2,3,4$	...	...		...
		$l = 1$	$R_{1,1,18,1}$	$R_{1,2,18,1}$	...	$R_{1,42,18,1}$
	$k = 18$	$l = 2$	$R_{1,1,18,2}$	$R_{1,2,18,2}$	...	$R_{1,42,18,2}$
		$l = 3$	$R_{1,1,18,3}$	$R_{1,2,18,3}$	...	$R_{1,42,18,3}$
Subtotal for Asphalts		69				
Group II Heavy residual fuel oils $i = 2$	$k = 1$	$l = 1$	$R_{2,1,1,1}$	$R_{2,2,1,1}$	...	$R_{2,42,1,1}$
		$l = 2$	$R_{2,1,1,2}$	$R_{2,2,1,2}$	...	$R_{2,42,1,2}$
		$l = 3$	$R_{2,1,1,3}$	$R_{2,2,1,3}$	...	$R_{2,42,1,3}$
	$k = 20$	$l = 1,2,3$				
Subtotal for heavy residual fuel oils		77				
Total, two groups	38	146				

R_{ijkl}

i = 1, (Asphalts), i = 2, (Heavy residual fuel oils)
j = 1,2, . . . , 42 (Ratio identification)
k = 1,2, . . . , 18 or 20 (Sample number identification)
l = 1,2,3 or 4 or 5 (Non-uniform replicate identification)

Table III. Ratios, Sample Means, and Standard Deviations Used in the Standardization

Variable indices	Ratios, cm^{-1}/cm^{-1}	Sample means over both groups $\bar{R}_j$	Sample std dev over both groups s_j	j-Number	Variable indices	Ratios, cm^{-1}/cm^{-1}	Sample means over both groups $\bar{R}_j$	Sample std dev over both groups s_j	J-Number
1	720/ 810	0.9011	0.4417	1	22	810/ 720	1.5605	1.0913	18
2	720/ 870	0.9282	0.4402	2	23	870/ 720	1.4492	0.9325	19
3	720/1027	2.0848	1.1422	..	24	1027/ 720	0.6627	0.3977	20
4	720/1375	0.1673	0.0483	..	25	1375/ 720	6.5536	2.1227	21
5	720/1460	0.0668	0.0202	..	26	1460/ 720	16.4365	5.1924	22
6	720/1600	0.4790	0.1913	3	27	1600/ 720	2.5047	1.2217	23
7	810/ 870	1.0709	0.2267	4	28	870/ 810	0.9746	0.2014	24
8	810/1027	2.4697	0.8863	5	29	1027/ 810	0.4752	0.2326	25
9	810/1375	0.2400	0.1411	6	30	1375/ 810	5.5035	2.5259	..
10	810/1460	0.0920	0.0468	7	31	1460/ 810	13.3866	5.4041	26
11	810/1600	0.5965	0.1914	8	32	1600/ 810	1.8530	0.5763	27
12	870/1027	2.3682	0.8840	9	33	1027/ 870	0.5035	0.2750	28
13	870/1375	0.2293	0.1274	..	34	1375/ 870	5.8899	3.1589	29
14	870/1460	0.0881	0.0427	10	35	1460/ 870	14.2983	6.9191	..
15	870/1600	0.5633	0.1608	11	36	1600/ 870	1.9241	0.5508	30
16	1027/1375	0.1023	0.0515	12	37	1375/1027	12.7261	6.6201	31
17	1027/1460	0.0399	0.0186	13	38	1460/1027	31.3061	15.3465	32
18	1027/1600	0.2636	0.0985	14	39	1600/1027	4.3411	1.7111	33
19	1375/1460	0.4013	0.0479	15	40	1460/1375	2.5261	0.2961	34
20	1375/1600	2.9257	1.0648	16	41	1600/1375	0.3867	0.1351	35
21	1460/1600	7.2261	2.3874	17	42	1600/1460	0.1514	0.0435	36

This program performed a multiple discriminant analysis in a stepwise manner. At each step, one variable was entered into the set of discriminating variables. The variable having the largest F was selected for inclusion. A minimum F value for inclusion (F to enter) of 0.01, a reasonable criterion, was placed in the input to the program.

Execution of the stepwise discriminant analysis resulted in the inclusion of 36 of the 42 variables. The 42 variable indices and the 36 selected j-numbers are shown in Table III. The analysis provided the b's for Equations 1 and 2, respectively, as shown in Table IV.

The foregoing data transformations and evaluations of the discriminant functions are illustrated by the following example: let the length from base to peak for wavenumber 720 cm^{-1} be 0.70 cm. Let the length from peak to bottom of the spectrum at $A = \infty$ be 6.80 cm. Using a copy of the original infrared spectrum, Figure 1, the values at the other wavenumbers are, for 810 cm^{-1} = 1.75 and 6.00; for 870 cm^{-1} = 1.80 and 6.80; for 1027 cm^{-1} = 1.10 and 8.35; for 1375 cm^{-1} = 2.95 and 4.60; for 1460 cm^{-1} = 6.20 and 2.30; and for 1600 cm^{-1} = 2.35 and 6.95, all in centimeters.

Table IV. Coefficients for Both Discriminant Functions When Using Standardized Variables in Equations 1 and 2

(1)	Asphalts	(2)	Heavy residual fuel oils	(1)	Asphalts	(2)	Heavy residual fuel oils
b_0	−7.60987	b_0'	−6.24480	b_{19}	−43.55099	b_{19}'	39.02415
b_1	−25.58461	b_1'	22.90611	b_{20}	4.07508	b_{20}'	−3.64347
b_2	26.62715	b_2'	−23.81479	b_{21}	−29.84641	b_{21}'	26.78763
b_3	−11.64139	b_3'	10.41510	b_{22}	11.77883	b_{22}'	−10.59161
b_4	−36.14394	b_4'	32.34564	b_{23}	−7.65198	b_{23}'	6.85141
b_5	−1.98233	b_5'	1.75529	b_{24}	39.78601	b_{24}'	−35.56705
b_6	−4.09560	b_6'	3.87530	b_{25}	−9.39279	b_{25}'	8.42572
b_7	−68.58247	b_7'	61.14029	b_{26}	46.71219	b_{26}'	−41.73645
b_8	62.78983	b_8'	−56.07736	b_{27}	−74.16132	b_{27}'	66.26096
b_9	5.81115	b_9'	−5.20976	b_{28}	−9.84835	b_{28}'	8.81127
b_{10}	62.29706	b_{10}'	−55.72067	b_{29}	−47.45398	b_{29}'	42.41000
b_{11}	−67.18605	b_{11}'	60.05365	b_{30}	72.33095	b_{30}'	−64.69716
b_{12}	−33.92418	b_{12}'	30.28230	b_{31}	9.34523	b_{31}'	−8.37640
b_{13}	22.85368	b_{13}'	−20.40642	b_{32}	−2.56282	b_{32}'	2.32982
b_{14}	24.45306	b_{14}'	−21.88516	b_{33}	−7.29110	b_{33}'	6.53285
b_{15}	8.81806	b_{15}'	−7.90065	b_{34}	3.05424	b_{34}'	−3.03080
b_{16}	25.18460	b_{16}'	−22.33339	b_{35}	11.34435	b_{35}'	−10.42913
b_{17}	−24.20792	b_{17}'	21.43787	b_{36}	−1.50103	b_{36}'	1.60187
b_{18}	63.63580	b_{18}'	−57.02377				

The absorbances were calculated from the above seven pairs of values and were: at 720 cm^{-1} = 0.0426; at 810 cm^{-1} = 0.1112; at 870 cm^{-1} = 0.1020; at 1027 cm^{-1} = 0.0537; at 1375 cm^{-1} = 0.2152; at 1460 cm^{-1} = 0.5677; and at 1600 cm^{-1} = 0.1265.

At this point computer assistance was necessary because of the large numbers of calculations required in obtaining the 42 possible absorbance ratios, their standardization, and the deletion of the standardized ratios excluded by the stepwise discriminant analysis.

Table V shows the wavenumber ratios in the order generated, the 42 absorbance ratios, the 42 standardized absorbance ratios, and the 36 j-numbers utilized by the stepwise discriminant analysis (from the list of included variates). The 36 standardized ratios are substituted in Equations 1 and 2 where the coefficients for these equations are found in Table IV. The manner of calculation and the results are shown in Table VI. Since the z_o function calculates to be a greater value than the z_a function, −3.157 *vs.* −11.079, the unknown is classified as a heavy residual fuel oil. For each observation, ratios of absorbances are formed and then standardized, and both discrimi-

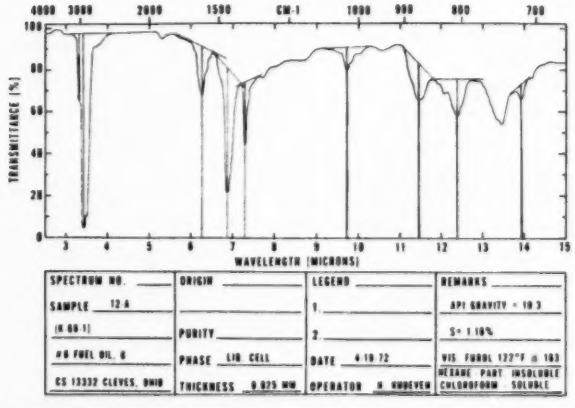

Figure 1. Infrared spectrum of a heavy residual fuel oil, No. 6, having the following properties: Sulfur, 1.16%; API gravity, 10.3; viscosity furol at 122 °F, 163. Source, Mid-Continent, summer production

Table V. Example of the Calculation of Absorbance Ratios, Standardization of Absorbance Ratios, and the Selection of the 36 Included Ratios

Wavenumbers, cm^{-1}/cm^{-1}	Absorbance ratios	Standardized absorbance ratios	J-Number
720/ 810	0.3816	−1.1762	1
720/ 870	0.4172	−1.1609	2
720/1027	0.7917	−1.1321	..
720/1375	0.1977	0.6295	..
720/1460	0.0750	0.4024	..
720/1600	0.3364	−0.7456	3
810/ 870	1.0933	0.0987	4
810/1027	2.0747	−0.4457	5
810/1375	0.5182	1.9718	6
810/1460	0.1964	2.2322	7
810/1600	0.8815	1.4894	8
870/1027	1.8977	−0.5323	9
870/1375	0.4740	1.9214	..
870/1460	0.1797	2.1457	10
870/1600	0.8063	1.5113	11
1027/1375	0.2498	2.8633	12
1027/1460	0.0947	2.9506	13
1027/1600	0.4249	1.6365	14
1375/1460	0.3791	−0.4655	15
1375/1600	1.7011	−1.1501	16
1460/1600	4.4877	−1.1470	17
810/ 720	2.6204	0.9712	18
870/ 720	2.3968	1.0163	19
1027/ 720	1.2630	1.5093	20
1375/ 720	5.0571	−0.7050	21
1460/ 720	13.3411	−0.5961	22
1600/ 720	2.9728	0.3831	23
870/ 810	0.9147	−0.2975	24
1027/ 810	0.4820	0.0290	25
1375/ 810	1.9299	−1.4148	..
1460/ 810	5.0912	−1.5350	26
1600/ 810	1.1345	−1.2467	27
1027/ 870	0.5270	0.0852	28
1375/ 870	2.1099	−1.1966	29
1460/ 870	5.5661	−1.2620	..
1600/ 870	1.2403	−1.2416	30
1375/1027	4.0039	−1.3175	31
1460/1027	10.5627	−1.3517	32
1600/1027	2.3537	−1.1615	33
1460/1375	2.6381	0.3783	34
1600/1375	0.5878	1.4887	35
1600/1460	0.2228	1.6418	36

Table VI. Example for the Calculation of Discriminant Functions

Term	Equation 1 b_j's	x_i's		Equation 2 b_j''s	x_i's	
Constant	−7.60987		+	−6.24480		+
1	(−25.58461)	(−1.1762)	+	(+22.90611)	(−1.1762)	+
2	(+26.62715)	(−1.1609)	+	(−23.81479)	(−1.1609)	+
3	(−11.64139)	(−0.7456)	+	(+10.41510)	(−0.7456)	+
4	(−36.14394)	(+0.0987)	+	(+32.34564)	(+0.0987)	+
5	(−1.98233)	(−0.4457)	+	(+1.75529)	(−0.4457)	+
6	(−4.09560)	(+1.9718)	+	(+3.87530)	(+1.9718)	+
7	(−68.58247)	(+2.2322)	+	(+61.14029)	(+2.2322)	+
8	(+62.78983)	(+1.4894)	+	(−56.07736)	(+1.4894)	+
9	(+5.81115)	(−0.5323)	+	(−5.20976)	(−0.5323)	+
10	(+62.29706)	(+2.1457)	+	(−55.72067)	(+2.1457)	+
11	(−67.18605)	(+1.5113)	+	(+60.05365)	(+1.5113)	+
12	(−33.92418)	(+2.8633)	+	(+30.28230)	(+2.8633)	+
13	(+22.85368)	(+2.9506)	+	(−20.40642)	(+2.9506)	+
14	(+24.45306)	(+1.6365)	+	(−21.88516)	(+1.6365)	+
15	(+8.81806)	(−0.4655)	+	(−7.90065)	(−0.4655)	+
16	(+25.18460)	(−1.1501)	+	(−22.33339)	(−1.1501)	+
17	(−24.20792)	(−1.1470)	+	(+21.43787)	(−1.1470)	+
18	(+63.63580)	(+0.9712)	+	(−57.02377)	(+0.9712)	+
19	(−43.55099)	(+1.0163)	+	(+39.02415)	(+1.0163)	+
20	(+4.07508)	(+1.5003)	+	(−3.64347)	(+1.5093)	+
21	(−29.84641)	(−0.7050)	+	(+26.78763)	(−0.7050)	+
22	(+11.77883)	(−0.5961)	+	(−10.59161)	(−0.5961)	+
23	(−7.65198)	(+0.3831)	+	(+6.85141)	(+0.3831)	+
24	(+39.78601)	(−0.2975)	+	(−35.56705)	(−0.2975)	+
25	(−9.39279)	(+0.0290)	+	(−8.42572)	(+0.0290)	+
26	(+46.71219)	(−1.5350)	+	(−41.73645)	(−1.5350)	+
27	(−74.16132)	(−1.2467)	+	(+66.26096)	(−1.2467)	+
28	(−9.84835)	(+0.0852)	+	(+8.81127)	(+0.0852)	+
29	(−47.45398)	(−1.1966)	+	(+42.41000)	(−1.1966)	+
30	(+72.33095)	(−1.2416)	+	(−64.69716)	(−1.2416)	+
31	(+9.34523)	(−1.3175)	+	(−8.37640)	(−1.3175)	+
32	(−2.56282)	(−1.3517)	+	(+2.32982)	(−1.3517)	+
33	(−7.29110)	(−1.1615)	+	(+6.53285)	(−1.1615)	+
34	(+3.54245)	(+0.3783)	+	(−3.03080)	(+0.3783)	+
35	(+11.34435)	(+1.4887)	+	(−10.42193)	(+1.4887)	+
36	(−1.50103)	(+1.6418)		(+1.60187)	(+1.6418)	
	z_a = −11.079056			z_b = −3.156959		

Table VII. Parameters of Standardized Data

Variable, cm^{-1}/cm^{-1}	Means: Asphalts	Means: Heavy residual fuel oils	Means: Over both groups	Standard deviations: Asphalts	Standard deviations: Heavy residual fuel oils	Variable cm^{-1}/cm^{-1}	Means: Asphalts	Means: Heavy residual fuel oils	Means: Over both groups	Standard deviations: Asphalts	Standard deviations: Heavy residual fuel oils
720/ 810	0.756	−0.678	0.000	0.723	0.675	810/ 720	−0.633	0.567	0.000	0.223	1.083
720/ 870	0.462	−0.414	0.000	0.714	1.041	870/ 720	−0.533	0.477	0.000	0.282	1.161
720/1027	0.354	−0.317	0.000	0.996	0.897	1027/ 720	−0.302	0.270	0.000	0.977	0.947
720/1375	0.439	−0.393	0.000	0.853	0.962	1375/ 720	−0.460	0.413	0.000	0.616	1.097
720/1460	0.579	−0.519	0.000	0.909	0.769	1460/ 720	−0.573	0.514	0.000	0.626	0.995
720/1600	0.354	−0.317	0.000	0.771	1.077	1600/ 720	−0.445	0.399	0.000	0.423	1.186
810/ 870	−0.634	0.568	0.000	0.719	0.869	870/ 810	0.672	−0.602	0.000	0.916	0.617
810/1027	−0.548	0.491	0.000	0.877	0.838	1027/ 810	0.509	−0.456	0.000	1.183	0.458
810/1375	−0.564	0.505	0.000	0.363	1.115	1375/ 810	0.476	−0.426	0.000	0.666	1.059
810/1460	−0.552	0.494	0.000	0.416	1.109	1460/ 810	0.495	−0.444	0.000	0.683	1.034
810/1600	−0.763	0.683	0.000	0.464	0.845	1600/ 810	0.773	−0.693	0.000	0.721	0.644
870/1027	−0.127	0.114	0.000	1.232	0.722	1027/ 870	0.242	−0.216	0.000	1.278	0.589
870/1375	−0.401	0.359	0.000	0.494	1.189	1375/ 870	0.110	−0.098	0.000	0.633	1.236
870/1460	−0.347	0.311	0.000	0.557	1.193	1460/ 870	0.081	−0.073	0.000	0.605	1.253
870/1600	−0.440	0.394	0.000	0.569	1.134	1600/ 870	0.321	−0.287	0.000	0.720	1.126
1027/1375	−0.152	0.136	0.000	0.927	1.048	1375/1027	0.089	−0.080	0.000	0.948	1.044
1027/1460	−0.057	0.051	0.000	1.028	0.978	1460/1027	0.030	−0.027	0.000	0.935	1.060
1027/1600	0.035	−0.031	0.000	1.282	0.659	1600/1027	0.146	−0.131	0.000	1.234	0.714
1375/1460	0.338	−0.302	0.000	0.924	0.974	1460/1375	−0.357	0.320	0.000	0.836	1.031
1375/1600	0.023	−0.021	0.000	0.668	1.228	1600/1375	−0.209	0.187	0.000	0.633	1.214
1460/1600	−0.068	0.061	0.000	0.623	1.246	1600/1460	−0.074	0.066	0.000	0.754	1.179

Table VIII. Results of Discriminant Function Analysis—146 Samples

Asphalts			Heavy residual fuel oils			Asphalts			Heavy residual fuel oils		
A	B	C	A	B	C	A	B	C	A	B	C
1	a	1.000	1	o	1.000	40	a	1.000	40	o	1.000
2	a	1.000	2	o	1.000	41	a	1.000	41	o	1.000
3	a	1.000	3	o	1.000	42	a	1.000	42	o	1.000
4	a	1.000	4	o	1.000	43	a	1.000	43	o	1.000
5	a	1.000	5	o	1.000	44	a	1.000	44	o	1.000
6	a	1.000	6	o	1.000	45	a	1.000	45	o	1.000
7	a	1.000	7	o	1.000	46	a	1.000	46	o	1.000
8	a	1.000	8	o	1.000	47	a	1.000	47	o	1.000
9	a	1.000	9	o	1.000	48	a	1.000	48	o	1.000
10	a	1.000	10	o	1.000	49	a	1.000	49	o	1.000
11	a	1.000	11	o	1.000	50	a	1.000	50	o	1.000
12	a	1.000	12	o	1.000	51	a	1.000	51	o	1.000
13	a	1.000	13	o	1.000	52	a	1.000	52	o	1.000
14	a	1.000	14	o	1.000	53	a	1.000	53	o	1.000
15	a	1.000	15	o	1.000	54	a	1.000	54	o	1.000
16	a	1.000	16	o	1.000	55	a	1.000	55	o	1.000
17	a	1.000	17	o	1.000	56	a	1.000	56	o	1.000
18	a	1.000	18	o	1.000	57	a	1.000	57	o	1.000
19	a	1.000	19	o	1.000	58	a	1.000	58	o	1.000
20	a	1.000	20	o	0.999	59	a	1.000	59	o	1.000
21	a	1.000	21	o	1.000	60	a	1.000	60	o	1.000
22	a	1.000	22	o	1.000	61	a	1.000	61	o	1.000
23	a	1.000	23	o	1.000	62	a	1.000	62	o	1.000
24	a	1.000	24	o	1.000	63	a	1.000	63	o	1.000
25	a	1.000	25	o	1.000	64	a	1.000	64	o	1.000
26	a	1.000	26	o	1.000	65	a	1.000	65	o	1.000
27	a	1.000	27	o	1.000	66	a	1.000	66	o	1.000
28	a	1.000	28	o	1.000	67	a	1.000	67	o	1.000
29	a	1.000	29	o	1.000	68	a	1.000	68	o	1.000
30	a	1.000	30	o	1.000	69	a	1.000	69	o	1.000
31	a	0.999	31	o	1.000				70	o	1.000
32	a	1.000	32	o	1.000				71	o	1.000
33	a	1.000	33	o	1.000				72	o	1.000
34	a	0.895	34	o	1.000				73	o	1.000
35	a	1.000	35	o	1.000				74	o	1.000
36	a	1.000	36	o	1.000				75	o	1.000
37	a	1.000	37	o	1.000				76	o	1.000
38	a	1.000	38	o	1.000				77	o	1.000
39	a	1.000	39	o	1.000						

[a] Column heading definitions: A, observation number; B, classified as a = asphalt, o = heavy residual fuel oil; C, probability for correct classification.

Table IX. Average Probability of Correctly Classifying 146 Samples

Groups	Average probability of correct classification
Asphalts	0.999
Heavy residual fuel oils	0.999
Average over both groups	0.999

nant functions are evaluated. The higher algebraic value, whether of z_o or of z_a, will determine the classification of the given replicate.

Since the data have been standardized, all variables have approximately the same mean and standard deviation, Table VII.

Therefore, the magnitude of the coefficients, b_j's and b_j''s, for the variables contained in Table IV can be considered as a measure of discriminatory power for each ratio or variable. The five most important variables in decreasing order of importance are 1600 cm^{-1}/810 cm^{-1},

Table X. Classification of Observations

Probability of correct classification	Numbers of observations: By group: Asphalts	Heavy residual fuel oils	Both groups, total	Per cent	Cumulative per cent for both groups: From top	From bottom
1.000	67	76	143	97.9	97.9	100.0
0.999 to less than 1.000	1	1	2	1.4	99.3	2.1
0.900 to less than 0.999	0	0	0	0.0	99.3	0.7
0.800 to less than 0.900	1	0	1	0.7	100.0	0.7
Total	69	77	146	100.0		

Table XI. Characterization of Unknown Samples Using Linear Discriminant Function Analysis

Wavenumber ratios, cm^{-1}/cm^{-1}	J-Value	Sample 1		Sample 2	
		Absorbance ratios	Standardized absorbance ratios	Absorbance ratios	Standardized absorbance ratios
720/ 810	1	1.2645	0.8227	1.1895	0.6531
720/ 870	2	0.9325	0.0096	0.7995	−0.2925
720/1600	3	0.3503	−0.6726	0.3311	−0.7732
810/ 870	4	0.7374	−1.4707	0.6721	−1.7588
810/1027	5	2.0568	−0.4659	2.3750	−0.1069
810/1375	6	0.1506	−0.6339	0.1478	−0.6543
810/1460	7	0.0591	−0.7057	0.0643	−0.5933
810/1600	8	0.2771	−1.6692	0.2784	−1.6625
870/1027	9	2.7891	0.4762	3.5338	1.3186
870/1460	10	0.0801	−0.1881	0.0957	0.1775
870/1600	11	0.3757	−1.1664	0.4142	−0.9273
1027/1375	12	0.0732	−0.5644	0.0622	−0.7784
1027/1460	13	0.0287	−0.6043	0.0271	−0.6925
1027/1600	14	0.1347	−1.3085	0.1172	−1.4862
1375/1460	15	0.3921	−0.1940	0.4352	0.7078
1375/1600	16	1.8394	−1.0202	1.8839	−0.9784
1460/1600	17	4.6917	−1.0616	4.3286	−1.2137
810/ 720	18	0.7908	−0.7053	0.8407	−0.6596
870/ 720	19	1.0724	−0.4041	1.2508	−0.2128
1027/ 720	20	0.3845	−0.6995	0.3540	−0.7763
1375/ 720	21	5.2503	−0.6140	5.6895	−0.4070
1460/ 720	22	13.3918	−0.5864	13.0727	−0.6478
1600/ 720	23	2.8543	0.2861	3.0201	0.4218
870/ 810	24	1.3561	1.8942	1.4879	2.5487
1027/ 810	25	0.4862	0.0471	0.4210	−0.2330
1460/ 810	26	16.9335	0.6563	15.5505	0.4004
1600/ 810	27	3.6092	3.0472	3.5925	3.0182
1027/ 870	28	0.3585	−0.5272	0.2830	−0.8019
1375/ 870	29	4.8957	−0.3147	4.5487	−0.4246
1600/ 870	30	2.6616	1.3389	2.4145	0.8903
1375/1027	31	13.6547	0.1403	16.0740	0.5057
1460/1027	32	34.8285	0.2295	36.9327	0.3666
1600/1027	33	7.4234	1.8013	8.5323	2.4494
1460/1375	34	2.5507	0.0830	2.2977	−0.7714
1600/1375	35	0.5436	1.1617	0.5308	1.0667
1600/1460	36	0.2131	1.4192	0.2310	1.8300

1600 cm^{-1}/870 cm^{-1}, 810 cm^{-1}/1460 cm^{-1}, 870 cm^{-1}/1600 cm^{-1}, and 810 cm^{-1}/720 cm^{-1}. The five least important variables in decreasing order are 1027 cm^{-1}/720 cm^{-1}, 1460 cm^{-1}/1375 cm^{-1}, 1460 cm^{-1}/1027 cm^{-1}, 810 cm^{-1}/1027 cm^{-1}, and 1600 cm^{-1}/1460 cm^{-1}. The *j*-numbers are ranked in descending order of the absolute magnitude of the coefficients. These *j*-numbers are: 27, 30, 07, 11, 18, 08, 10, 29, 26, 19, 24, 04, 12, 21, 02, 01, 16, 14, 17, 13, 22, 03, 35, 28, 25, 31, 15, 23, 33, 09, 06, 20, 34, 32, 05, 36. The wavenumber ratios corresponding to these *j*-values are found in Table V.

The evaluation by both discriminant functions of the 146 observations and resulting probability for correct classification are shown in Table VIII.

Sample one, replicate one in the asphalt group has a probability of 1.000 of being an asphalt and probability of 1.000 − 1.000 = 0.000 of being an oil and was assigned to the asphalt group. By discriminant analysis, all of the replicates were correctly classified as an asphalt or a heavy residual fuel oil. Table IX gives the average probability of correct classification for each group taken separately and both groups collectively.

From the last line of that table, it can be seen that the odds of correct classification are 0.999 to 0.001 or about 999 to 1.

Some interpretation of the frequency distribution of the probabilities shown in Table X follows: 97.9% of the 146 observations had a probability of correct classification equal to 1.000. From the last column, 2.1% had a probability of less than 1.000, or, in the next to last column, 99.3% had a probability of 0.999 or greater.

Several asphalts and No. 6 residual fuel oils that were not included in the original 146 replicates were tested using this procedure. The ratios of absorbances were standardized and the discriminant functions evaluated. Results revealed that asphalts and No. 6 residual fuel oils not taken from the original 146 observations, but representative of the two groups, can be correctly classified with a high degree of certainty.

Table XI shows the data on tested materials associated with a spill believed to be asphalt that was obtained on the Ohio River near Pittsburgh, Pa. For sample 1, from the spill itself, the two function values are: z_a = −4.68050, z_o = −8.92630, and the probabilities for correct classification are: Group a = 0.985, Group o = 0.015. For sample 2, a suspected source asphalt, the two function values are: z_a = −5.22735, z_o = −8.40503, and the probabilities for correct classification are: Group a = 0.960, Group o = 0.040. The two samples were correctly classified with a high degree of probability. From the results of this study, the potential for detailed identification of environmental samples is recognized.

CONCLUSION

The objectives mentioned in this paper have been realized in that the observations on 18 asphalt samples and on 20 heavy residual fuel oil samples, for the first time, have been correctly classified. Discriminant function analysis is

a technique for utilizing more of the information which has been generated in the laboratory. Thus, a combination of infrared measurements, data treatment and transformation, and discriminant function analysis through computer assistance has resulted in effecting a more precise and accurate method of distinguishing between these heavy petroleum materials. An established statistical technique has hereby been applied successfully to a recently developed infrared procedure for heavier petroleum products to provide a useful and powerful technique for classification and may result in a promising application to further identification.

ACKNOWLEDGMENT

It is a special pleasure to acknowledge the support and interest given by D. G. Ballinger, Director of the Analytical Quality Control Laboratory, throughout the investigation. Much credit is due R. C. Kroner, Chief, Physical and Chemical Methods, for his interest and help in this study. The assistance given by Hilda Knueven is greatly appreciated.

Received for review January 4, 1973. Accepted August 16, 1973. Mention of the products and manufacturers is for identification purposes only and does not imply endorsement by the Environmental Protection Agency.